APPLICATION OF MACHINE LEARNING METHODS FOR ANALYZING DATA FROM THE NOMENCLATURE DIRECTORY OF THE ENTERPRISE RESOURCE PLANNING SYSTEM, PART 1

Mushtak O.I.[1], Limanovskaya O.V.[1], Lebedev A.S.[1,2,3*]

[1)] Ural Federal University, Yekaterinburg, Russia
[2)] Ural Mining and Metallurgical Company (UMMC), Yekaterinburg, Russia
[3)] None-state Higher Educational Establishment "UMMC Technical University", Yekaterinburg, Russia
*E-mail: aslebedev@urfu.ru

Applying of enterprise resource planning systems (ERP - Enterprise resource planning) as an integral part of corporate information systems of an enterprise has become the most important factor ensuring the success of a modern company in a rapidly changing market for goods and services.

The need to implement the ERP system arose in the Ural Mining and Metallurgical Company (UMMC) – the largest producer of copper, zinc, coal and precious metals in Russia. Since 2008, the UMMC has been actively implementing at its enterprises a single corporate information system (UMMC CIS) based on SAP ERP.

The adopted system uses standard reference information. One of the main elements of the standard reference information (SRI) of the UMMC CIS is the material master record - the main source of information about individual materials in an enterprise.

The unified directory of materials in the UMMC CIS contains more than 500 000 elements and constantly is replenished and improved by own service on work with SRI.

A software product is being developed for the development of a corporate directory of materials in the UMMC CIS. The developed product should allow automating a number of functions that are currently performed with considerable laboriousness or require a long data processing time.

The main objectives of the developed product:

1. Comparison of the reference book of the UMMC organization materials and all-corporate reference materials book;

a. The developed product should allow quickly correlating two data arrays available in tables UMMC CIS or in the form of the tables SAP ERP and the external file of a certain structure. One of the files (arrays) will be the directory of materials of the UMMC organization, and the second array, available in the SAP ERP tables, is the corporate directory of materials.

The task of the developed functionality is to improve the quality of searching for records of the general corporate directory of materials corresponding to the records in the directories of materials of UMMC organizations.

2. The automated classification of materials;

a. Using methods of machine learning and methods of data processing, the developed functionality should automatically to assign classification attributes to records of the all-corporate reference book of materials from UMMC CIS.

3. Fast assessment of stocks when purchasing;
4. Identification of illiquid and unclaimed stocks.

Export data in CSV-format from corporate directory of materials in the UMMC CIS was used as basic data.

Preparation of data for the analysis was carried out by the morphological pymorphy2 analyzer and the CountVectorizer module of sklearn library.

The analysis of data consisted in multiclass classification. The following methods were used: random forest, naive Bayes and XGBoosting.

The results are shown in Fig.1.

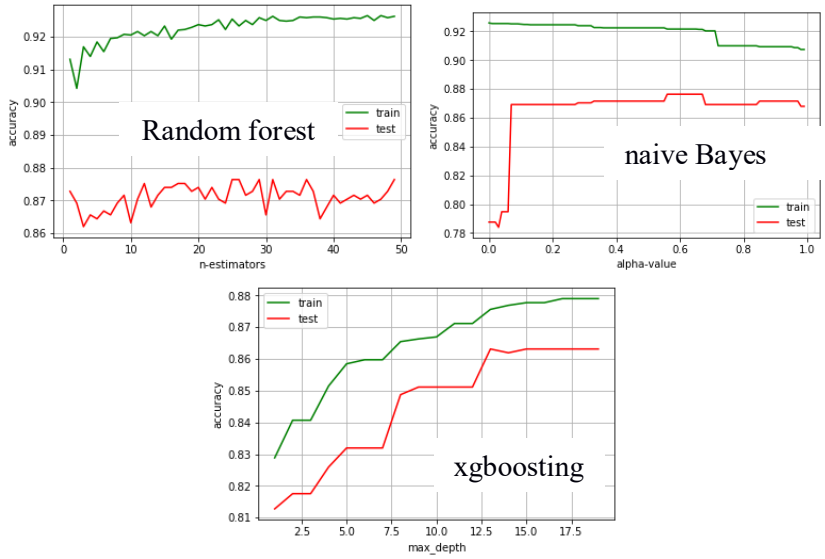

Fig. 1. Results of applying different methods

The study of the dependence of the quality of the model on the parameters of the classifiers is presented in the second part.